Molecular Plant Advance Access published August 8, 2008

Molecular Plant • Pages 1–14, 2008

Integration of Cytological Features with Molecular and Epigenetic Properties of Rice Chromosome 4

Ben-Liang Yin[a], Lan Guo[a,b], Dong-Fen Zhang[c], William Terzaghi[d,e], Xiang-Feng Wang[a,b,d], Ting-Ting Liu[a], Hang He[a], Zhu-Kuan Cheng[c] and **Xing Wang Deng[a,d,1]**

a National Institute of Biological Sciences, Beijing 102206, China
b Peking–Yale Joint Research Center for Plant Molecular and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China
c Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
d Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520–8104, USA
e Department of Biology, Wilkes University, Wilkes-Barre, PA 18766, USA

ABSTRACT **It has been reported that rice chromosome 4 has eight major heterochromatic knobs within the heterochromatic half and that this organization correlates with chromosomal-level transcriptional activity. To better understand this chromosomal organization, we created a model based on the statistical distribution of various types of gene models to divide chromosome 4 into 17 euchromatic and heterochromatic regions that correspond with the cytological staining. Fluorescence in-situ hybridization (FISH) experiments using a set of bacterial artificial chromosome (BAC) clones from chromosome 4 placed all 18 clones in the region predicted by the model. Elevated levels of H3K4 di- and tri-methylation detected by chromatin-immunoprecipitation (ChIP) on chip were correlated with euchromatic regions whereas lower levels of these two modifications were detected in heterochromatic regions. Small RNAs were more abundant in the heterochromatic regions. To validate these findings, H3K4 trimethylation, H3K9 acetylation, H4K12 acetylation, and H3K9 di- and tri-methylation of 19 individual genes were measured by ChIP–PCR. Genes in heterochromatic regions had elevated H3K9 di- and tri-methylation while genes in euchromatic regions had elevated levels of the other three modifications. We also assayed cytosine methylation of these genes using the restriction enzymes McrBC, HapII, and Msp I. This analysis indicated that cytosines of transposable elements and some genes located in heterochromatic regions were methylated while cytosines of the other genes were unmethylated. These results suggest that local transcriptional activity may reflect the organization of the corresponding part of the chromosome. They also indicate that epigenetic regulation plays an important role in correlating chromosomal organization with transcriptional activity.**

Key words: **chromosome organization; FISH; histone modification; DNA methylation.**

INTRODUCTION

Rice (*Oryza sativa*) is one of the most important cereal crops and is widely studied as a model monocotyledenous plant (Harushima et al., 1998; Shimamoto and Kyozuka, 2002). The rice genome has been almost completely sequenced, including difficult regions such as centromeres (Cheng et al., 2001b; Nonomura and Kurata, 2001; Feng et al., 2002; Sasaki et al., 2002; Rice Chromosome 10 Sequencing Consortium, 2003; Wu et al., 2004; Zhang et al., 2004; International Rice Genome Sequencing Project, 2005). This research has culminated in the recent release of Version 5 of the rice genome annotation by the Institute of Genome Research (www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml). This provides a very powerful resource for studying rice and all plants.

Now that the rice genome has been sequenced, it is time to determine its function. Computational analysis has been a powerful tool for identifying functional genomic elements; however, computational predictions require experimental validation (Yuan et al., 2003; Ouyang et al., 2007). Collection and sequencing of cDNAs and ESTs are an important way to validate the results of computational annotation, since this verifies the existence of gene models predicted by computer

[1] To whom correspondence should be addressed. E-mail xingwang.deng@yale.edu, fax 203–432–5726.

doi: 10.1093/mp/ssn037

modeling (Wu et al., 2002; Kikuchi et al., 2003). Recently, tiling-path microarrays have been developed as additional tools to study transcriptional activity at the genomic level (Shoemaker et al., 2001; Yamada et al., 2003; Bertone et al., 2005; Jiao et al., 2005; Li et al., 2005). These chips detect transcription across the entire genome and also enable the study of other biological questions such as developmental regulation of transcription, histone modification, and DNA methylation (Huebert et al., 2006; Millar and Grunstein, 2006; Plass and Smiraglia, 2006; Sinha et al., 2006; Zhang et al., 2006; Oakes et al., 2007; Zilberman et al., 2007).

Genome-wide analyses show that eukaryotic chromosomes have complex structures that serve very important functions. In addition to forming the templates for DNA replication and RNA transcription, DNA sequences also determine the structure and function of chromosomes through their organization. Organization into euchromatic and heterochromatic regions profoundly affects the expression of sequences contained within these regions. Accordingly, it is imperative to understand how the division of chromosomes into euchromatic and heterochomatic regions is regulated. DNA sequence itself may carry information for chromosomal organization. For example, in *Drosophila*, some transposons have affinity for heterochromatic protein 1 (HP1), which plays a key role in heterochromatin formation (Fanti et al., 1998). Therefore, the completed sequence of the rice genome may help determine how cells define heterochromatin and euchromatin.

Heterochromatin was first identified as a chromatin that stained intensely throughout the cell cycle and was therefore assumed to be inactive (Heitz, 1928; Hennig, 1999). Recent studies in rice and *Arabidopsis* revealed that heterochromatin has dynamic staining characteristics and may therefore be more active than previously believed (Fransz et al., 1998; Cheng et al., 2001a; Lysak et al., 2006). Nonetheless, heterochromatin is linked with reduced transcriptional activity and recombinational capability (CSHL/WUGSC/PEB Arabidopsis Sequencing Consortium, 2000; Hoekenga et al., 2000; Stam et al., 2002; Mittelsten Scheid et al., 2003). Genomic research has shown that heterochromatin has large proportions of tandem repeats, silent transposon-related elements (TEs) and other non-coding DNA sequences (Lippman et al., 2004). Furthermore, some regulatory epigenetic modifications such as DNA methylation and histone modifications have been shown to help form heterochromatin (Bender, 2004; Tariq and Paszkowski, 2004). For example, methylation of H3K9 is a very important mark of heterochromatin. In contrast, methylation of lysine residue 4 on Histone H3 (H3K4) is thought to be a mark of euchromatin that promotes transcription. Acetylation of histone lysines such as H3K9 and H4K12 is also linked with euchromatic regions and gene activation (Houben et al., 1997; Grant-Downton and Dickinson, 2005; Benhamed et al., 2006). Histone modifications have been widely studied in *Arabidopsis* but relatively less in rice, so these results need confirmation in rice.

Another factor that has been proposed to play an important role in directing histone and DNA modifications is RNA interference (Bernstein and Allis, 2005; Gendrel and Colot, 2005; Wassenegger, 2005). Small RNA molecules, including those which mediate RNA interference, are often missed in genome projects. Small interfering RNA (siRNA) of 21–24 nucleotides have been shown to target the complementary mRNA for degradation and are often involved in gene silencing (Filipowicz et al., 2005). However, other small RNAs of similar size function instead by targeting the complementary genomic DNA sequence for regulatory epigenetic modifications such as histone modifications and DNA methylation (Matzke and Birchler, 2005). Recently, a large number of rice small RNA sequences have been reported, providing a powerful resource for analyzing relationships between small RNAs and chromosomal organization (Nobuta et al., 2007).

Rice chromosome 4 has been intensively studied. Cytological staining shows that it has eight major heterochromatic knobs around the centromere and almost half of the chromosome is heterochromatic (Jiao et al., 2005). Recently, its centromere has been defined and completely sequenced (Cheng et al., 2001a; Zhang et al., 2004). We have used tiling-path microarrays to characterize transcriptional activity along the full length of chromosome 4 and found lower levels of transcription in the general vicinities of the heterochromatic knobs (Jiao et al., 2005). However, the exact boundaries of these heterochromatic regions and the ways in which their formation is regulated are still unclear.

In this study, we have improved the demarcation of these boundaries and started to address how the formation of heterochromatic regions is regulated. We first constructed a model that used differential distribution of transposons and expressed genes in heterochromatic and euchromatic regions to place the boundaries between these regions. We then classified the gene loci of chromosome 4 into four categories, depending upon whether they resembled transposons and whether there was experimental evidence for their expression, and used the distribution of these loci to subdivide chromosome 4 into 17 regions that corresponded well with the distribution of heterochromatic knobs based on cytological staining. We then showed that levels of transcription and of certain histone modifications and the abundance of small RNAs also varied between the euchromatic and heterochromatic regions defined by our model. These results were validated by studying histone modifications and DNA methylation of 19 individual genes. These results indicate that we can use this model to map the boundaries of heterochromatic and euchromatic regions on the remaining chromosomes and then use this information to improve our understanding of how the formation of heterochromatin is regulated.

RESULTS

Heterochromatic Regions Are Enriched in Transposable Elements and Are Correlated with Lower Levels of Transcription

TIGR Rice Genome Annotation Release Version 5 identifies 5372 gene loci on chromosome 4. We classified these loci as

TE or non-TE protein-coding genes (nTE), since TEs are important components of heterochromatin. We also divided them into groups with (SG) and without (UG) cDNA or EST experimental support. Since previous studies showed that heterochromatin is associated with lower levels of gene expression and silent TEs, we analyzed the distribution of these different types of loci along rice chromosome 4 by calculating the numbers of loci in a sliding 500-kb window with a step of 50 kb and then plotting their distribution.

The density of gene loci varies little along chromosome 4, maintaining an average of 75 loci per 500 kb (Figure 1B). However, there is a dramatic difference in the distribution of SG and UG, where UG are far more abundant in the first 18 Mb, while SG are more abundant in the remainder of the chromosome (Figure 1C). Cytological studies have shown that the first 18 Mb are highly heterochromatic (Jiao et al., 2005). Our analysis shows that there are more TE–UG and nTE–UG in the first 18 Mb, while there are more nTE–SG in the rest of chromosome 4 (Figure 1A). These data indicate that heterochromatin is closely correlated with the distribution of UG and lower transcriptional activity. To better visualize this difference, we plotted the distribution of nTE–UG and nTE–SG along chromosome 4 (Figure 1D). Both are rare in the first 18 Mb whereas silent TEs are relatively abundant. This suggests that we can use the distribution of different types of loci to define the borders of euchromatic and heterochromatic regions.

Division of Chromosome 4 into Euchromatin-Like and Heterochromatin-Like Regions

Previous studies have shown that rice chromosome 4 has eight major heterochromatic knobs (Jiao et al., 2005). We therefore divided the chromosome into eight heterochromatic, eight euchromatic, and one centromeric region, based on cytological staining. Their designations and approximate boundaries are shown in Figure 2A.

To better define the borders of the heterochromatic regions, we plotted the percentage of SG loci along chromosome 4. As shown in Figure 2B, on average, this percentage is much lower in the first 18 Mb than in the remainder of the chromosome. Closer examination of the first 18 Mb uncovered nine regions that have lower-than-average percentages of SG. Previous studies had located the core of chromosome 4's centromere at about 9.6 Mb (Zhang et al., 2004). Here, we observed a region with a low percentage of SG located around 9.6 Mb; therefore, we defined this region as the centromere. On either side of the centromere, there are four regions with a lower percentage of SG that appear to correspond with the eight major heterochromatic knobs characterized by cytological staining. They also have higher levels of silent TE and reduced levels of SG (Figure 1), whereas the intervening regions have higher percentages of expressed genes. We therefore call the regions with lower SG percentages heterochromatin-like, and those with higher SG percentages euchromatin-like. Table 1 lists the borders of each region on chromosome 4.

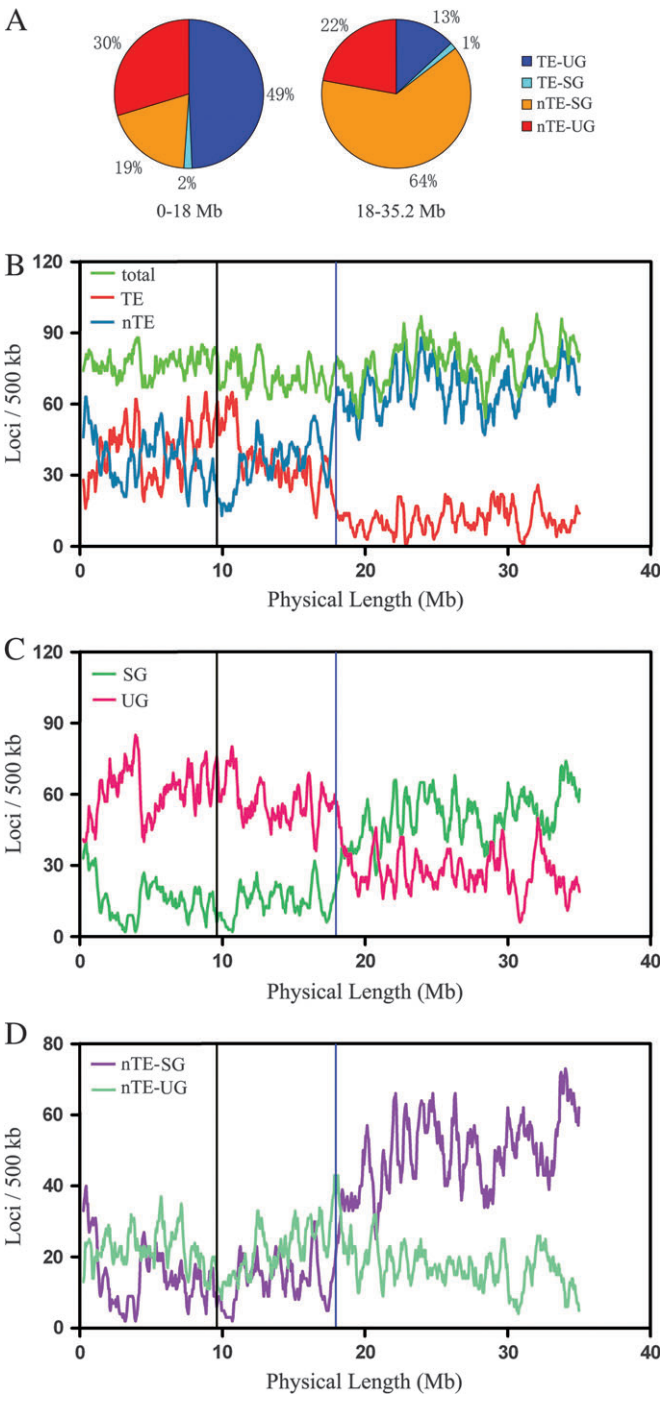

Figure 1. Heterochromatin Is Enriched in TEs and Correlated with Lower Levels of Gene Transcription in Rice Chromosome 4.

(A) Frequencies of different types of gene loci in the first 18 Mb and the remainder of chromosome 4.

(B) Distributions of TEs, nTEs, and total gene loci on chromosome 4.

(C) Distributions of SG and UG on chromosome 4.

(D) Distributions of nTE–SG and nTE–UG on chromosome 4.

In (B)–(D), the gray line shows the position of the centromere and the blue line indicates the border at about 18 Mb between the last major heterochromatic knob and the euchromatic region on the long arm of rice chromosome 4. Gene loci are counted in a sliding 500-kb window with a step of 50 kb.

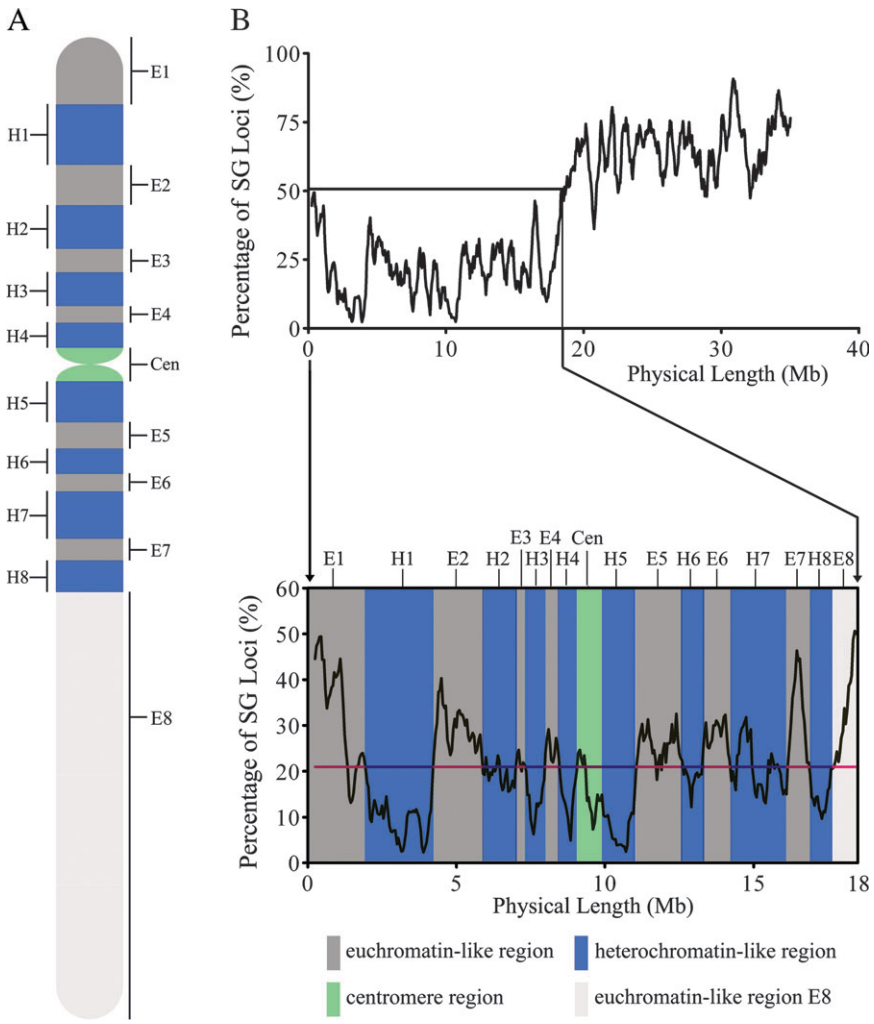

Figure 2. Division of Rice Chromosome 4 into Euchromatin-Like and Heterochromatin-Like Regions.

(A) Cartoon of rice chromosome 4 and designation of different regions. According to cytological staining, rice chromosome 4 can be divided into a centromere (Cen), eight euchromatic (E1–E8), and eight heterochromatic regions (H1–H8).

(B) Subdivision of rice chromosome 4. The percentage of SG loci is used to divide chromosome 4 into 17 regions. The red line shows the average percentage of SG loci in the first 18 Mb of chromosome 4. The regions whose percentage of SG loci is above the red line are classed as euchromatin-like, while those whose percentage of SG loci is under the red line are either heterochromatin-like or centromeric regions. The percentage is counted in a sliding 500-kb window with a step of 50 kb.

FISH Results Are Consistent with the Division of Rice Chromosome 4 into 17 Different Regions

To validate the division of rice chromosome 4, we performed fluorescence in-situ hybridization (FISH) on pachytene chromosomes with 18 BAC clones randomly selected from different parts of the first 18 Mb of chromosome 4 (Table 1). As shown in Figure 3, results of these experiments were consistent with the divisions shown in Figure 2. For example, clone OSJNBa27H09 clearly hybridized to the first euchromatic region near the first major heterochromatic knob (Figure 3M). Its coordinates are 1.70–1.86 Mb, whereas euchromatic region E1 extends from 0 to 1.95 Mb (Table 1). This result is thus consistent with the definition of the region E1. Results for the other BAC clones are also consistent with the region they are placed in according to their coordinates, thus validating the subdivision of rice chromosome 4 shown in Figure 2.

Microarray Data Show Correlations between Histone Modifications and Chromosomal Organization

We previously measured the RNA levels of rice transcriptionally active regions using a re-array (Li et al., 2007). We therefore used these data to study transcriptional activity within the various regions of chromosome 4 by calculating the average signal intensities of SG and UG in euchromatic and heterochromatic regions. Figure 4A shows that the average intensity of SG is higher than UG in both regions, and that intensity of both classes of loci is greater in euchromatic than in heterochromatic regions. These results also help validate the division of chromosome 4 into heterochromatic and euchromatic regions, since heterochromatin is correlated with reduced transcriptional activity in many organisms.

Histone modifications are widely studied as important epigenetic regulatory mechanisms and are found to be closely correlated with transcriptional activity. Among these, di- and tri-methylated H3K4 are associated with euchromatic regions and help activate gene transcription. We used chromatin-immunoprecipitation followed by microarray analysis (ChIP on chip) to identify regions of chromosome 4 that were enriched in di- and tri-methylated H3K4. We used the modification density (see Methods) to measure the level of histone modification. As shown in Figure 4B and 4C, the modification density of loci in euchromatic regions was higher than in heterochromatic regions.

Table 1. Chromosomal Subdivisions and Cytological Locations of Some BACs in Rice Chromosome 4.

Region	Start (Mb)	BAC	BAC coordinates (Mb)
E1	0.00	OSJNBa0027H09	1.70–1.86
H1	1.95	OSJNBa0070D17	2.17–2.34
		OSJNBa0056L23	2.52–2.7
		OSJNBa0089E12	3.67–3.82
		OSJNBa0032N05	4.03–4.19
E2	4.25	OSJNBa0039G19	5.04–5.18
H2	5.85	OSJNBa0045O17	5.87–6.01
		OSJNBa0059D20	6.12–6.27
		OSJNBA0014F04	6.74–6.91
E3	7.00	OSJNBa0035O13	7.15–7.3
H3	7.30	OSJNBA0021F22	7.50–7.68
E4	8.00	OSJNBA0036B17	8.31–8.41
H4	8.45	–	
Cen	9.10	–	
H5	9.95	OSJNBA0017P10	10.01–10.27
		OSJNBA0072D08	10.91–11.06
E5	11.10	OSJNBa0009K15	11.60–11.8
H6	12.60	OSJNBA0035B13	13.2–13.3
E6	13.30	–	
H7	14.20	OSJNBA0032I19	15.7–15.9
E7	16.15	–	
H8	16.85	OSJNBA0034E24	17.0–17.2
E8	17.70	–	

Moreover, both in euchromatic and heterochromatic regions, the modification density of UG loci is lower than SG in the putative promoter (the 1-kb region upstream of the annotated transcription start site) and gene body (the annotated transcribed region). These results show that di- and tri-methylated H3K4 are associated with high expression levels and euchromatic regions in rice.

We calculated the modification density of each region and did a cluster analysis to show the differences between regions (Figure 4D). Both types of histone modifications were more abundant in euchromatic regions, and were also more abundant in gene bodies than in promoters. These results further illustrate the differences between heterochromatic and euchromatic regions, and show that histone modifications are associated with gene expression and chromosomal organization.

Distribution of Small RNAs on Rice Chromosome 4

Rice small RNA transcripts have been studied using MPSS (Nobuta et al., 2007). We used rice MPSS data to analyze the distribution of small RNAs on chromosome 4 by counting the number of small RNAs in a 500-kb sliding window with a step of 50 kb (Figure 5). Small RNAs were mainly detected in the heterochromatic regions of chromosome 4, since the numbers in the first 18 Mb were much higher than in the remainder of the chromosome. However, small RNA distribution did not show the clear differences between heterochromatic and euchromatic regions observed with gene loci in Figure 2. Therefore, although it is intriguing that the frequency of small RNAs is substantially higher in the heterochromatic arm of chromosome 4, more study will be needed to determine whether this correlates directly with chromatin structure.

Histone Modifications of Individual Genes on Rice Chromosome 4

All of the above analyses were performed at a whole-chromosome level, so they needed validation at the level of individual genes. We therefore randomly chose a set of 19 genes located in different regions of chromosome 4 to compare the relative enrichment of histone modifications. Annotations, regions they are assigned to, and classifications of these genes are listed in Table 2. We selected two representative tissues, roots and shoots of 7-day-old light-grown rice seedlings, for chromatin immunoprecipitation followed by real-time PCR. Previous studies showed that some histone modifications were found on the histones binding promoter regions (Narlikar et al., 2002; Millar and Grunstein, 2006), so we analyzed the 2 kb surrounding the start codons and found a peak of GC content about 200 bp downstream of the ATG (Supplemental Figure 1A). Moreover, nTE–SG have higher GC content than the other classes (Supplemental Figure 1B), which suggests that this pattern may be related to the gene expression level. Accordingly, we designed the primers for real-time PCR to bind near the ATG of their respective genes, except for the primers for gene 19, which bind in an exon and were selected as the endogenous control. Sequences of the corresponding primers are listed in Supplemental Table 2. We divided these genes into the five classes shown in Figure 6, indicated by differently colored backgrounds according to their locus type and chromosomal region.

Five types of histone modifications were measured. Trimethylated H3K4, acetylated H3K9, and acetylated H4K12 are reported to be associated with euchromatin and activation of gene expression, whereas di- and tri-methylated H3K9 are thought to be associated with heterochromatin and gene suppression in other species. Histone modifications showed a close association with gene expression and chromosomal organization (Figure 6). In all regions, the genes without experimental support had lower relative enrichments than the ones with experimental support for the three activating histone modifications, while the repressive histone modifications showed the opposite pattern. These results show that the three activating histone modifications were associated with higher levels of transcriptional activity in both euchromatic and heterochromatic regions, whereas the two suppressive modifications were associated with reduced transcriptional activity. Moreover, in general, the relative enrichments of activating histone modifications were much higher in euchromatic than in heterochromatic regions and the suppressive histone modifications were much lower in euchromatic regions. These results indicate that in rice chromosome 4, histone modifications were associated with chromosomal organization.

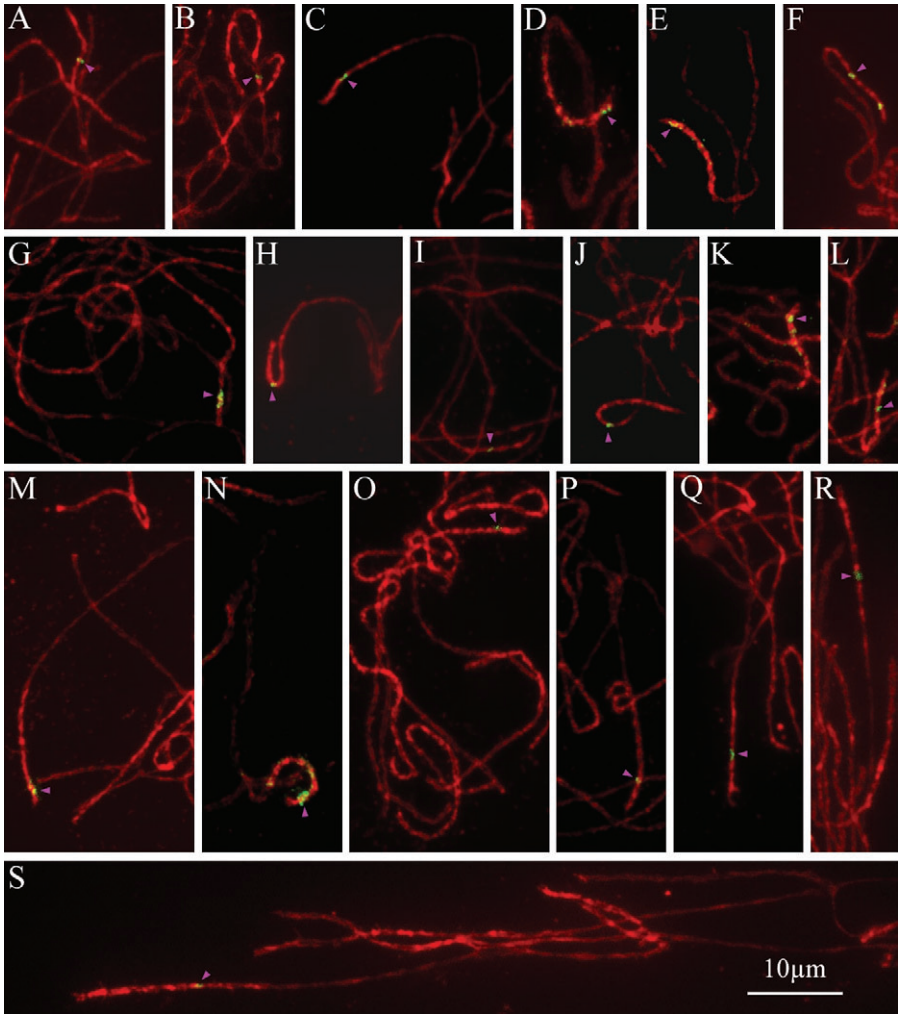

Figure 3. Cytological Localization of 18 BACs by Fluorescence In-Situ Hybridization.

BACs labeled with Cy3 were used to probe rice pachytene chromosomes and detected by green fluorescence after irradiation with 532 nm light. In each image, the pink arrow shows the location of the hybridization signal. BACs used in these experiments are: **(A)** OSJNBa0017P10; **(B)** OSJNBa0034E24; **(C)** and **(Q)** OSJNBa0036B17; **(D)** OSJNBa0032N05; **(E)** OSJNBa0070D17; **(F)** OSJNBa0072D08; **(G)** OSJNBa0014F04; **(H)** OSJNBa0035O13; **(I)** OSJNBa0021F22; **(J)** OSJNBa0032I19; **(K)** OSJNBa0056L23; **(L)** OSJNBa0059D20; **(M)** OSJNBa0027H09; **(N)** OSJNBa0089E12; **(O)** OSJNBa0039G19; **(P)** OSJNBa0045O17; **(R)** OSJNBa0009K15; **(S)** OSJNBa0035B13.

These results also show that in rice, some histone modifications are correlated with each other. The pattern of trimethylated H3K4 is similar to the other two activating modifications, acetylated H3K9 and acetylated H4K12. Moreover, the patterns of the two suppressive methylations of H3K9 are also similar to each other. The histone modifications of the tested genes show similar patterns in the selected tissues. This is consistent with the finding that the transcriptional activities of these genes are also similar in the selected tissues (Supplemental Figure 2).

TEs and nTEs Have Different DNA Methylation Patterns

DNA methylation is another important mechanism for epigenetic regulation. Previous studies revealed that DNA methylation is also associated with gene suppression and may be correlated with histone modifications such as dimethylated H3K9 (Jackson et al., 2002; Johnson et al., 2002). To further study the correlation between these epigenetic modifications, we assayed DNA methylation along rice chromosome 4 by McrBC and Hap II/Msp I digestion. McrBC is an endonuclease that binds the methylated half-sites (G/A)^{m}C and cleaves between them (Lippman et al., 2004). Hap II and Msp I are isoschizomers that cut CCGG sites, except Hap II digestion is blocked by CG and CNG methylation whereas Msp I digestion is only blocked by CNG methylation (Johnson et al., 2002). We therefore digested rice genomic DNA with these enzymes, then used PCR to assay the methylation status of the 19 genes listed in Table 2. Hap II/Msp I digestion only detected methylation of genes 1, 2, and 3, while McrBC only detected methylation of gene 2 (Figure 7A). Genes 1, 2, and 3 do not have experimental support and are located in heterochromatic regions with highly methylated H3K9 (Figure 6). This indicates that in rice, DNA methylation is associated with suppression of nTEs.

Transposable elements are an important component of the rice genome. Most fall into two classes—the retroelements (class I), whose mobility must be mediated via a reverse transcribed RNA, and DNA transposons (class II), which often encode a transposase and possess terminal inverted-repeats (Turcotte et al., 2001). Most transposable elements are silenced and dispersed in the heterochromatic regions. Since one of the main functions of DNA methylation is silencing transposable elements, we studied the methylation patterns of each class

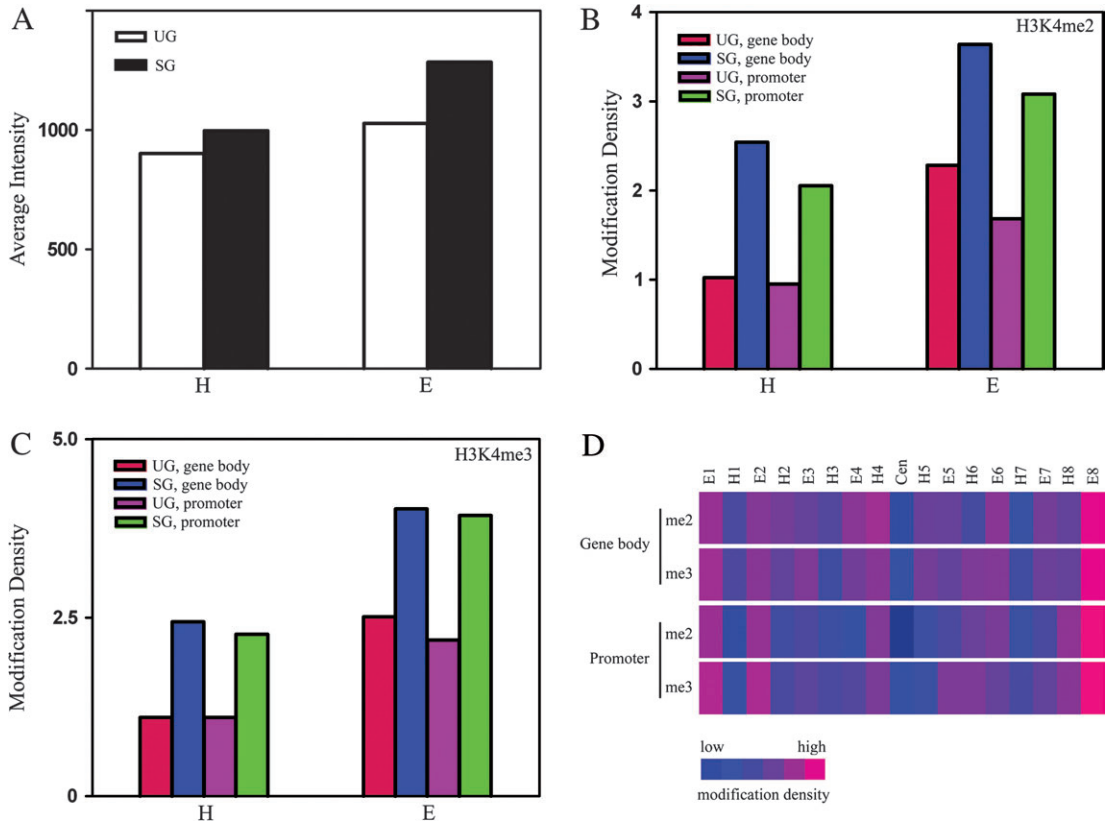

Figure 4. Microarray Data Show Differences between Euchromatic and Heterochromatic Regions in Gene Expression and Histone Modifications.

(A) Average microarray hybridization intensities of SG and UG loci in euchromatic and heterochromatic regions.
(B) H3K4me2 and **(C)** H3K4me3 modification of the four types of gene loci as measured by ChIP on chip.
(D) Modification densities of different regions of chromosome 4.
H, heterochromatin; E, euchromatin. The modification density is the number of probes binding modified targets per Kb.

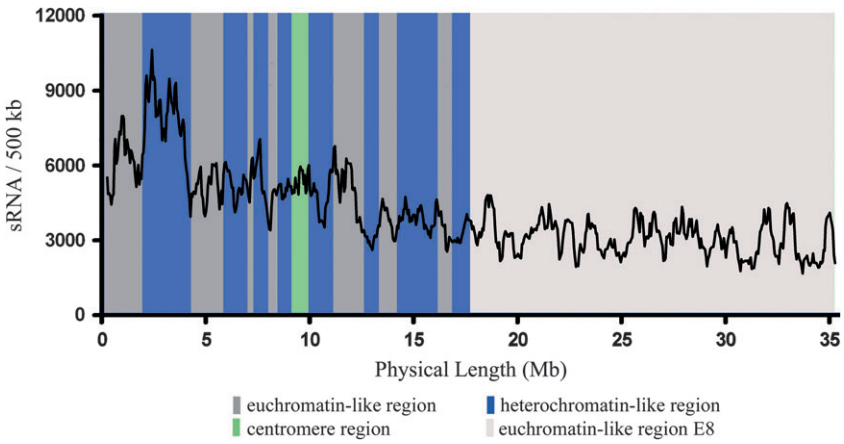

Figure 5. The Distribution of Small RNAs on Rice Chromosome 4.

The number of small RNAs is counted in a sliding 500-kb window with a step of 50 kb. The different chromosomal regions are represented by various colors.

of TE. The results showed that all class I TEs were methylated at CCGG sites in upstream, downstream, and gene body regions (Figure 7B). McrBC assays showed that sequences flanking most retroelements (class I) were methylated (Figure 7B), whereas only half of sequences flanking DNA transposons (class II) were methylated (Figure 7C). These results suggest that DNA methylation may play a different role in the two classes of TEs.

Sequences Surrounding the Selected Genes

Upstream and downstream factors may influence the histone modifications and transcription of all of our tested genes.

Table 2. Loci Selected to Design Primers for ChIP–PCR and DNA Methylation.

No.	Locus number	Region	Locus type	Annotation
1	Os04g10890	H2	nTE–UG	Zinc knuckle family protein
2	Os04g13270	H3	nTE–UG	Hypothetical protein
3	Os04g22450	H6	nTE–UG	Hypothetical protein
4	Os04g22660	H6	nTE–SG	Tropinone reductase 2, putative
5	Os04g11130	H2	nTE–SG	Low-molecular-weight cysteine-rich protein LCR68 precursor, putative
6	Os04g12720	E3	nTE–SG	Indole-3-acetate beta-glucosyltransferase, putative
7	Os04g22220	E5	nTE–SG	Pathogenesis-related protein PRB1-3 precursor, putative
8	Os04g13170	E3	nTE–SG	Ribosomal RNA apurinic site-specific lyase, putative
9	Os04g13210	E3	nTE–SG	Multidrug resistance-associated protein 4, putative
10	Os04g43080	E8	nTE–UG	Hypothetical protein
11	Os04g45060	E8	nTE–UG	MYB14, putative
12	Os04g34710	E8	nTE–UG	Hypothetical protein
13	Os04g53040	E8	TE–SG	Retrotransposon protein, putative, unclassified
14	Os04g48400	E8	nTE–SG	Protein HOTHEAD precursor, putative
15	Os04g35560	E8	nTE–SG	Tetrachloro-P-hydroquinone reductive dehalogenase, putative
16	Os04g48010	E8	nTE–SG	Signal transducer, putative
17	Os04g48050	E8	nTE–SG	Protein binding protein, putative
18	Os04g41580	E8	nTE–SG	Expressed protein
19	Os04g48360	E8	nTE–UG	Hypothetical protein

Previous studies showed that DNA methylation and histone modifications can function over very large scales encompassing as much as 2 Mb (Frigola et al., 2006; Clark, 2007). To examine the effect of adjacent transcriptional activity, we calculated the percentage of expressed loci in a 50-kb window surrounding each of the 19 tested genes (Figure 8A). More expressed genes were found within 50 kb of the tested genes located in euchromatic regions than those in heterochromatic regions. For detailed investigation of the surrounding genes, all loci in the 50-kb window are shown (Figure 8B). nTE–SG located in euchromatic regions have many similar neighbors, whereas nTE–UG are surrounded by TEs. Modifications to individual genes may therefore reflect much larger-scale modifications to the surrounding region.

DISCUSSION

Heterochromatic regions have long been recognized as large chromosomal regions that stain more intensely than euchromatic regions, while, more recently, they have been shown to have reduced transcriptional and recombinational activity (Kagawa et al., 2002; Ebert et al., 2006; Grewal and Jia, 2007). However, although heterochromatic regions are readily identified by cytological staining, determining their precise boundaries at the molecular level has rarely been addressed. Here, we describe a statistical model that subdivides rice chromosome 4 into heterochromatic and euchromatic regions based on the distribution of various types of gene loci that is consistent with cytological staining of pachytene chromosome 4.

It has been found previously that heterochromatin is correlated with gene suppression and heterochromatic regions are enriched in silenced TEs and tandem repeats (Martienssen, 2003; Lippman et al., 2004). We therefore hypothesized that we could use the combined distribution of TEs and expressed gene loci to identify heterochromatic regions. To develop our model, we used rice chromosome 4, since it has been extensively studied at both the molecular and the cytological levels. We chose cDNA or EST support as the criterion for identifying expressed loci because this provides experimental validation that the locus is transcribed. However, alternative methods such as tiling-path microarrays gave similar results (Figure 4; Li et al., 2008). We therefore believe that this is a useful way to map the distribution of expressed loci that can be applied to many eukaryotic genome projects.

Distributions of TEs and expressed loci vary substantially along chromosome 4 (Figure 1), permitting the subdvision into euchromatic and heterochromatic regions shown in Figure 2. The FISH experiments using BAC clones provide experimental support for these chromosomal subdivisions (Figure 3). The distributions of histone modifications shown in Figure 4D also show substantial variation coinciding with these chromosomal subdivisions. In addition, the detailed examinations of 19 individual genes show that each one is expressed and modified in a manner consistent with the region to which it is assigned (Figures 5–7). Moreover, neighboring genes are also expressed and modified in a manner consistent with the region to which they are assigned (Figure 8). We are therefore confident that this model accurately assigns loci to euchromatic or heterochromatic regions. It will therefore be useful for predicting the

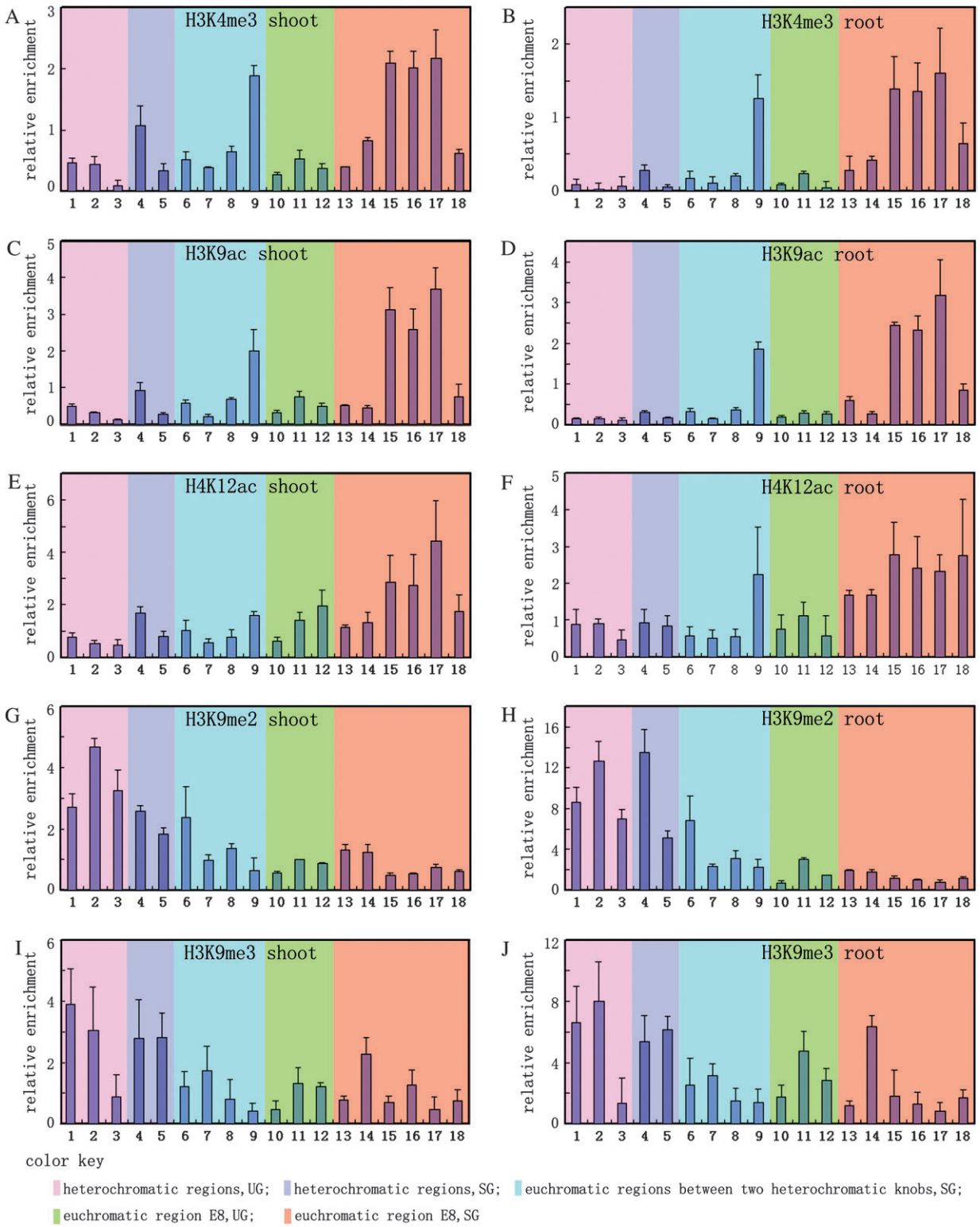

Figure 6. Histone Modifications of Individual Genes Show Region-Specific Features.

Histone modification levels were measured by ChIP followed by real-time PCR. Eighteen genes were selected to represent the various gene types and regions, which are distinguished by different colors. In **(A)**, **(C)**, **(E)**, **(G)**, and **(I)**, the tissue is 7-day-old rice shoots and in **(B)**, **(D)**, **(F)**, **(H)**, and **(J)**, the tissue is 7-day-old rice roots.

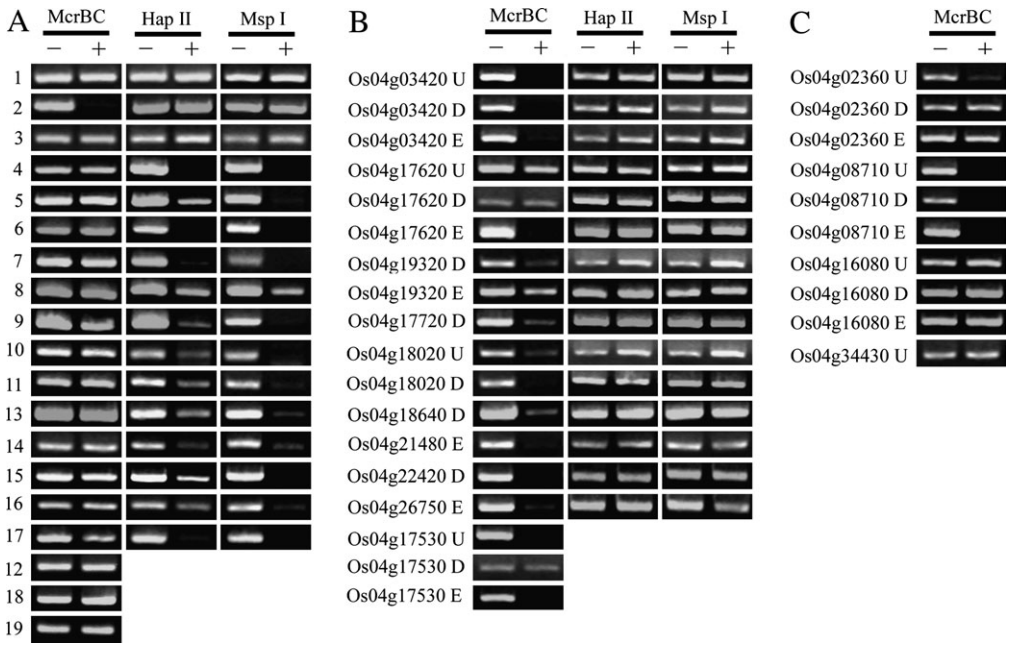

Figure 7. DNA Methylation of the Selected Genes and Some TEs.

DNA methylation patterns were assayed by digesting genomic DNA with McrBC, Hap II, and Msp I, then testing for digestion by PCR followed by gel electrophoresis.

(A) DNA methylation patterns of the genes tested for histone modifications.

(B) DNA methylation patterns of some class I TEs.

(C) DNA methylation patterns of some class II TEs.

The lanes have no Hap II and Msp I results because these amplicons have no CCGG sites. In (B) and (C), 'U' and 'D' indicate that the amplicon is within 1 kb upstream of the start codon and 1 kb downstream of the stop codon, respectively. 'E' indicates that the amplicon is within an exon of the gene.

structure of other eukaryotic chromosomes and the likely expression level of any gene based on its chromosomal location.

Our data also show close relationships of transcriptional activity, histone modifications and chromosome structure. Heterochromatic regions are closely correlated with methylated H3K9 and enrichment in TEs and protein-coding genes without cDNA or EST support. This is consistent with related studies in other eukaryotic species that show that chromatin structure mediated by histone modifications plays an important role in gene expression (Jackson et al., 2002; Tariq and Paszkowski, 2004; Vaillant and Paszkowski, 2007). Our data also support the observation that large regions are modified in similar manners, since the 19 genes that were studied intensively were all surrounded by sequences expressed in similar ways (Figure 8).

The ability to accurately map the borders between heterochromatic and euchromatic regions is crucial for understanding how eukaryotes decide which DNA sequences will be packaged into heterochromatin after DNA replication. Biochemical studies have shown that heterochromatin contains protein complexes such as heterochromatin protein 1 and enzymes for histone modifications (Narlikar et al., 2002). However, specific DNA sequences that mark heterochromatic regions have not been identified. Our model provides a way to help identify heterochromatin-specific sequences.

Our distribution analysis (Figure 1) shows that TEs and unexpressed genes were more abundant in the heterochromatic regions, while Figure 8 shows that unexpressed genes were surrounded by TEs or other unexpressed genes. We therefore hypothesize that the accumulation of TEs plays a very important role in large-scale silencing. TEs have relatively specific sequences that may be molecular markers recruiting specific protein complexes such as enzymes functioning in epigenetic regulation to bind and then trigger the modifications, resulting in the conversion of that region into heterochromatin. Under this hypothesis, entire regions with high concentrations of TE will be converted to heterochromatin, resulting in the silencing of non-TE genes contained in that region.

In conclusion, this study provides a powerful tool for helping to decipher one of the major challenges in genome biology—how cells choose to package DNA into heterochromatin or euchromatin.

METHODS

Classification and Distribution of Gene Loci on Rice Chromosome 4

Rice chromosome 4 annotation data were downloaded from the TIGR rice genome database (Version 5, January, 2007).

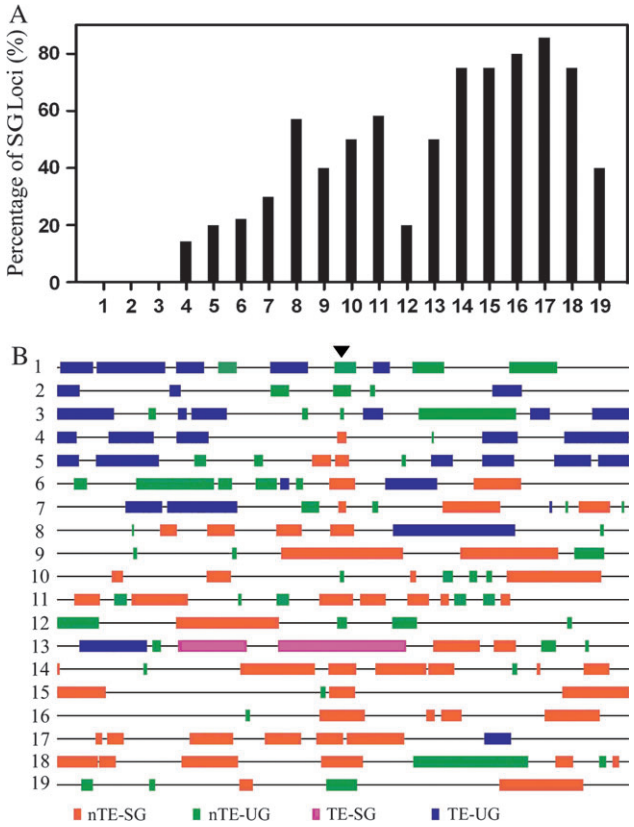

Figure 8. Features of Regions Surrounding the 19 Tested Gene Loci.
The numbers1 to 19 refer to the genes assayed in the histone modification and DNA methylation experiments.
(A) The percentage of SG loci in the 50-kb region surrounding the 19 tested gene loci.
(B) Detailed distributions of gene loci in the surrounding regions. The black triangles show the positions of the tested gene loci. The colors distinguish the gene types.

The table chr04.TU_model.brief.info.5 listed 6237 computationally predicted gene models. Because alternative splicing may result in several different gene models annotated at the same locus, we chose the locus rather than the predicted gene model as the parameter to count the distribution of transcriptional units. After combining gene models predicted at the same locus, we counted 5372 gene loci.

We classified the loci as either transposable elements (TE), or other protein-coding genes (nTE), according to the annotation of each locus. To link with gene expression, we further classed each group into those with (SG) or without experimental support (UG), resulting in four classes of gene loci.

We counted the number of each class of gene loci in a sliding 500-Kb window, plotting the numbers with a 50-Kb step to show the distributions of these loci along chromosome 4. We chose the percentage of SG within this sliding window as a quantitative parameter to show the distribution of gene expression along chromosome 4.

Plant Materials and Antibodies

The rice strain used in ChIP and DNA methylation assays was *japonica cv Nipponbare.* Seeds were imbibed at 37°C for 3 d to break dormancy and then cultured in water at 28°C in 16 h/8 h (day/night) light. Entire 7-day-old seedlings were collected to extract genomic DNA. Shoots and roots were collected separately as representative tissues for chromatin immunoprecipitation. Panicles for fluorescence in-situ hybridization were collected from field-grown plants. Five antibodies were used for chromatin immunoprecipitation. Anti-H3K9 acetylation and anti-H4K12 acetylation antibodies were described previously (Guo et al., 2006). Anti-H3K4 trimethylation (#05–745), anti-H3K9 di- (#07–441), and tri-methylation (#07–442) antibodies were purchased from Upstate Biotechnology.

Fluorescence In-Situ Hybridization Analysis

Fluorescence in-situ hybridization (FISH) was performed as described previously (Cheng et al., 2001c).

Microarray Data Analysis

The expression profile by a re-array was described previously (Li et al., 2007). We extracted the seedling shoot data and counted the average intensities of each class of gene in euchromatic and heterochromatic regions.

The ChIP on chip experiments and data analysis were previously described (Li et al., 2008). We used the data on di- and trimethylated H3K4 from light-grown shoots. The modification density was calculated as the average number of probes binding modified targets per kb. We then calculated the average modification densities for the different regions and different types of gene loci.

Chromatin Immunoprecipitation and Real-Time PCR

Chromatin immunoprecipitation was performed as described previously (Gendrel et al., 2005). Quantitative real-time PCR was performed using an ABI PRISM 7500 Fast Real-Time PCR System to measure the relative enrichment of modified histone-associated sequences. PCR was performed in triplicate using SYBR Premix Ex Taq Kit and run at 95°C for 2 min, 45 cycles of 95°C for 15 s, 60°C for 30 s, 72°C for 34 s. The PCR cycle threshold (CT) was first set automatically and then manually adjusted such that each signal was slightly above baseline. We used the comparative CT (ddCT) method to calculate the relative enrichment, using the input as the calibrator and primer set 19 as the endogenous control.
dCT = CTinterest – CTendo, and ddCT = dCTenrich – dCTinput. The final relative enrichment was 2^{-ddCT}.

Extraction of Rice Genomic DNA

Rice genomic DNA was extracted with CTAB. Freshly harvested seedlings were pulverized in liquid nitrogen, thawed in CTAB solution (2% CTAB, 0.1 M Tris-HCl, 0.02 M EDTA, 1.4 M NaCl, 0.2% β-Mercaptoethanol, 1 mg ml^{-1} Proteinase K, pH 8.0), then incubated at 50°C for 1 h with occasional agitation. After centrifugation at 8000 rpm for 10 min to pellet debris,

supernatants were transferred to new tubes and extracted with equal volumes of phenol-chloroform (1:1, v/v), then precipitated with two volumes of cold ethanol. The pellet was washed with 70% ethanol, air-dried, then dissolved in 100 μl water and digested with RNase. After RNase digestion, the DNA was extracted with an equal volume of phenol-chloroform, then precipitated with two volumes of cold ethanol. The pellet was washed with 70% ethanol, air-dried, then dissolved in water for use.

McrBC Digestion

McrBC digestion was performed according to the manufacturer's instructions. Genomic DNA (final concentration of 15 ng μl^{-1}) was digested with100 units McrBC in a volume of 100 μl for 5 h at 37°C. Water was substituted for enzyme as a negative control. PCR was performed with gene-specific primers, then gel electrophoresis was used to visualize the amplification results.

Hap II/Msp I Digestion

Genomic DNA was digested by Hap II and Msp I according to the manufacturer's instructions. Genomic DNA (final concentration of 15 ng μl^{-1}) was digested with100 units Hap II in 100 μl or 50 units Msp I in 100 μl for 5 h at 37°C. Water was substituted for enzyme as a negative control. PCR was performed with gene-specific primers, then gel electrophoresis was used to visualize the amplification results.

Primers

All primers used in ChIP–PCR and DNA methylation assays are listed in Supplemental Table 1 and Supplemental Table 2.

SUPPLEMENTARY DATA

Supplementary Data are available at www.mplant.oxfordjournals.org.

FUNDING

This work was supported by the grant from National '863' High-Tech Project of National Ministry of Science and Technology (2003AA 210070) and The 863 Rice Functional Genomics Program of National Ministry of Science and Technology, People's Republic of China.

ACKNOWLEDGMENTS

We thank many lab members for providing helpful comments during this work. No conflict of interest declared.

REFERENCES

Bender, J. (2004). Chromatin-based silencing mechanisms. Curr. Opin. Plant Biol. **7,** 521–526.

Benhamed, M., Bertrand, C., Servet, C., and Zhou, D.X. (2006). Arabidopsis GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. Plant Cell. **18,** 2893–2903.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev. **19,** 1635–1655.

Bertone, P., Gerstein, M., and Snyder, M. (2005). Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. Chromosome Res. **13,** 259–274.

Cheng, Z., Buell, C.R., Wing, R.A., Gu, M., and Jiang, J. (2001a). Toward a cytological characterization of the rice genome. Genome Res. **11,** 2133–2141.

Cheng, Z., Presting, G.G., Buell, C.R., Wing, R.A., and Jiang, J. (2001b). High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics. **157,** 1749–1757.

Cheng, Z., Stupar, R.M., Gu, M., and Jiang, J. (2001c). A tandemly repeated DNA sequence is associated with both knob-like heterochromatin and a highly decondensed structure in the meiotic pachytene chromosomes of rice. Chromosoma. **110,** 24–31.

Clark, S.J. (2007). Action at a distance: epigenetic silencing of large chromosomal regions in carcinogenesis. Hum. Mol. Genet. **16 Spec. No 1,** R88–R95.

Ebert, A., Lein, S., Schotta, G., and Reuter, G. (2006). Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res. **14,** 377–392.

Fanti, L., Dorer, D.R., Berloco, M., Henikoff, S., and Pimpinelli, S. (1998). Heterochromatin protein 1 binds transgene arrays. Chromosoma. **107,** 286–292.

Feng, Q., et al. (2002). Sequence and analysis of rice chromosome 4. Nature. **420,** 316–320.

Filipowicz, W., Jaskiewicz, L., Kolb, F.A., and Pillai, R.S. (2005). Post-transcriptional gene silencing by siRNAs and miRNAs. Curr. Opin. Struct. Biol. **15,** 331–341.

Fransz, P., Armstrong, S., Alonso-Blanco, C., Fischer, T.C., Torres-Ruiz, R.A., and Jones, G. (1998). Cytogenetics for the model system Arabidopsis thaliana. Plant J. **13,** 867–876.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., and Clark, S.J. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet. **38,** 540–549.

Gendrel, A.V., and Colot, V. (2005). Arabidopsis epigenetics: when RNA meets chromatin. Curr. Opin. Plant Biol. **8,** 142–147.

Gendrel, A.V., Lippman, Z., Martienssen, R., and Colot, V. (2005). Profiling histone modification patterns in plants using genomic tiling microarrays. Nat. Methods. **2,** 213–218.

Grant-Downton, R.T., and Dickinson, H.G. (2005). Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. Ann. Bot. (Lond.). **96,** 1143–1164.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nat. Rev. Genet. **8,** 35–46.

Guo, L., Yin, B., Zhou, J., Li, X., and Deng, X.W. (2006). Development of rabbit monoclonal and polyclonal antibodies for detection of site-specific histone modifications and their application in analyzing overall modification levels. Cell Res. **16,** 519–527.

Harushima, Y., et al. (1998). A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics. **148,** 479–494.

Heitz, E. (1928). Das heterochromatin der moose. Jehrb. Wiss. Botanik. **69,** 762–818.

Hennig, W. (1999). Heterochromatin. Chromosoma. **108,** 1–9.

Hoekenga, O.A., Muszynski, M.G., and Cone, K.C. (2000). Developmental patterns of chromatin structure and DNA methylation responsible for epigenetic expression of a maize regulatory gene. Genetics. **155,** 1889–1902.

Houben, A., Belyaev, N.D., Leach, C.R., and Timmis, J.N. (1997). Differences of histone H4 acetylation and replication timing between A and B chromosomes of brachycome dichromosomatica. Chromosome Res. **5,** 233–237.

Huebert, D.J., Kamal, M., O'Donovan, A., and Bernstein, B.E. (2006). Genome-wide analysis of histone modifications by ChIP-on-chip. Methods. **40,** 365–369.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. Nature. **436,** 793–800.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. **416,** 556–560.

Jiao, Y., et al. (2005). A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. Plant Cell. **17,** 1641–1657.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks: DNA methylation and histone H3 lysine 9 methylation. Curr. Biol. **12,** 1360–1367.

Kagawa, N., Nagaki, K., and Tsujimoto, H. (2002). Tetrad-FISH analysis reveals recombination suppression by interstitial heterochromatin sequences in rye (Secale cereale). Mol. Genet. Genomics. **267,** 10–15.

Kikuchi, S., et al. (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science. **301,** 376–379.

Li, L., et al. (2007). Global identification and characterization of transcriptionally active regions in the rice genome. PLoS ONE. **2,** e294.

Li, L., Wang, X., Xia, M., Stolc, V., Su, N., Peng, Z., Li, S., Wang, J., and Deng, X.W. (2005). Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. Genome Biol. **6,** R52.

Li, X., et al. (2008). High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. Plant Cell. **20,** 259–276.

Lippman, Z., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature. **430,** 471–476.

Lysak, M., Fransz, P., and Schubert, I. (2006). Cytogenetic analyses of Arabidopsis. Methods Mol. Biol. **323,** 173–186.

Martienssen, R.A. (2003). Maintenance of heterochromatin by RNA interference of tandem repeats. Nat. Genet. **35,** 213–214.

Matzke, M.A., and Birchler, J.A. (2005). RNAi-mediated pathways in the nucleus. Nat. Rev. Genet. **6,** 24–35.

Millar, C.B., and Grunstein, M. (2006). Genome-wide patterns of histone modifications in yeast. Nat. Rev. Mol. Cell Biol. **7,** 657–666.

Mittelsten Scheid, O., Afsar, K., and Paszkowski, J. (2003). Formation of stable epialleles and their paramutation-like interaction in tetraploid Arabidopsis thaliana. Nat. Genet. **34,** 450–454.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell. **108,** 475–487.

Nobuta, K., et al. (2007). An expression atlas of rice mRNAs and small RNAs. Nat. Biotechnol. **25,** 473–477.

Nonomura, K., and Kurata, N. (2001). The centromere composition of multiple repetitive sequences on rice chromosome 5. Chromosoma. **110,** 284–291.

Oakes, C.C., La Salle, S., Smiraglia, D.J., Robaire, B., and Trasler, J.M. (2007). A unique configuration of genome-wide DNA methylation patterns in the testis. Proc. Natl Acad. Sci. U S A. **104,** 228–233.

Ouyang, S., et al. (2007). The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. **35,** D883–D887.

Plass, C., and Smiraglia, D.J. (2006). Genome-wide analysis of DNA methylation changes in human malignancies. Curr. Top. Microbiol. Immunol. **310,** 179–198.

Rice Chromosome 10 Sequencing Consortium (2003). In-depth view of structure, activity, and evolution of rice chromosome 10. Science. **300,** 1566–1569.

Sasaki, T., et al. (2002). The genome sequence and structure of rice chromosome 1. Nature. **420,** 312–316.

Shimamoto, K., and Kyozuka, J. (2002). Rice as a model for comparative genomics of plants. Annu. Rev. Plant Biol. **53,** 399–419.

Shoemaker, D.D., et al. (2001). Experimental annotation of the human genome using microarray technology. Nature. **409,** 922–927.

Sinha, I., Wiren, M., and Ekwall, K. (2006). Genome-wide patterns of histone modifications in fission yeast. Chromosome Res. **14,** 95–105.

Stam, M., Belele, C., Dorweiler, J.E., and Chandler, V.L. (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev. **16,** 1906–1918.

Tariq, M., and Paszkowski, J. (2004). DNA and histone methylation in plants. Trends Genet. **20,** 244–251.

The Cold Spring Harbor Laboratory, Washington University Genome Sequencing Center, and PE Biosystems Arabidopsis Sequencing Consortium (2000). The complete sequence of a heterochromatic island from a higher eukaryote. Cell. **100,** 377–386.

Turcotte, K., Srinivasan, S., and Bureau, T. (2001). Survey of transposable elements from rice genomic sequences. Plant J. **25,** 169–179.

Vaillant, I., and Paszkowski, J. (2007). Role of histone and DNA methylation in gene regulation. Curr. Opin. Plant Biol. **10,** 528–533.

Wassenegger, M. (2005). The role of the RNAi machinery in heterochromatin formation. Cell. **122,** 13–16.

Wu, J., et al. (2002). A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell. **14,** 525–535.

Wu, J., et al. (2004). Composition and structure of the centromeric region of rice chromosome 8. Plant Cell. **16,** 967–976.

Yamada, K., et al. (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. Science. **302,** 842–846.

Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J., and Buell, C.R. (2003). The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. Nucleic Acids Res. **31,** 229–233.

Zhang, X., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell. **126,** 1189–1201.

Zhang, Y., et al. (2004). Structural features of the rice chromosome 4 centromere. Nucleic Acids Res. **32,** 2023–2030.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. **39,** 61–69.